Chin. Ann. Math.
24B:3(2003),349-358.

MAXIMUM INFORMATION AND OPTIMUM ESTIMATING FUNCTION**

LIN Lu*

Abstract

In order to construct estimating functions in some parametric models, this paper introduces two classes of information matrices. Some necessary and sufficient conditions for the information matrices achieving their upper bounds are given. For the problem of estimating the median, some optimum estimating functions based on the information matrices are acquired. Under some regularity conditions, an approach to carrying out the best basis function is introduced. In nonlinear regression models, an optimum estimating function based on the information matrices is obtained. Some examples are given to illustrate the results. Finally, the concept of optimum estimating function and the methods of constructing optimum estimating function are developed in more general statistical models.

Keywords Quasi (pseudo) Fisher information, Estimating function, Quasi score, Nonlinear regression model, Median regression model
2000 MR Subject Classification 62B10, 62F10
Chinese Library Classification O212.1 **Document Code** A
Article ID 0252-9599(2003)03-0001-10

§1. Introduction and Definition

An unbiased estimating function $g(\theta, y)$ is defined to be a function of the data y and parameter θ having zero mean for all θ. One purpose of an estimating function is to produce an estimate $\hat{\theta}$ of the parameter from data y, the estimate being obtained as a root of the equation $g(\theta, y) = 0$. Consequently, if the parameter θ is p-dimensional, it is necessary at least that the range of g is p-dimensional with nonvanishing derivative matrix.

If the $n \times 1$ observation y has mean $\mu(\theta) = (\mu_1(\theta), \ldots, \mu_n(\theta))'$ and covariance matrix $\sigma^2 V(\theta)$, both being known functions of the p-dimensional parameter θ and $V(\theta)$ being a positive definite matrix, then, an unbiased estimating function is defined as

$$q(\theta, y) = \sigma^{-2}\{\dot{\mu}(\theta)\}'\{V(\theta)\}^{-1}e(\theta), \tag{1.1}$$

where $\dot{\mu}$ is an $n \times p$ matrix with components $\frac{\partial \mu_i}{\partial \theta_j}$, $\theta = (\theta_1, \ldots, \theta_p)'$, $\mathrm{rank}\{\dot{\mu}(\theta)\} = p$ and $e(\theta) = y - \mu(\theta)$. We call $q(\theta, y)$ the quasi score function. The quasi score function is a linear

Manuscript received January 31, 2002.

*Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, 300071, China.
E-mail: lnlu66@eyou.com.

**Project supported by the National Natural Science Foundation of China (No.10171051), and Youth Teacher Foundation of Nankai University.

unbiased estimating function based only on the first two moments of the observations. It is well known that the quasi score function is the optimum in the class of linear unbiased estimating functions $\{\sigma^{-2}H(\theta)e(\theta) : H \text{ is a } p \times n \text{ matrix}\}$ (McCullagh[10], Godambe and Heyde[2], Li and McCullagh[7]). It is easily verified that

$$i_{\theta_0} = \mathrm{Cov}_{\theta_0}(q) = -E_{\theta_0}\Big(\frac{\partial q}{\partial \theta}\Big) = \sigma^{-2}\dot{\mu}'(\theta_0)V^{-1}(\theta_0)\dot{\mu}(\theta_0), \tag{1.2}$$

where θ_0 denotes the true value of θ. This matrix plays the role of Fisher information exactly as in fully parametric inference and under the usual kinds of limiting conditions the asymptotic covariance matrix of quasi likelihood estimator $\hat{\theta}$ is $i_{\theta_0}^{-1}$. So, in this paper, we call i_θ the quasi Fisher information matrix (QFI) of quasi score function $q(\theta, y)$. Similarly, QFI of an arbitrary unbiased estimating function $g(\theta, y) \in \{\sigma^{-2}H(\theta)e(\theta) : H \text{ is a } p \times n \text{ matrix}\}$ is defined as

$$i_\theta(H) = \sigma^{-2}\dot{\mu}'(\theta)H'(\theta)(H(\theta)V(\theta)H'(\theta))^{-1}H(\theta)\dot{\mu}(\theta) \tag{1.3}$$

because the asymptotic covariance matrix of $\hat{\theta}_H$ is $i_\theta(H)$, where $\hat{\theta}_H$ is the root of the equation $g(\theta, y) = 0$ (McCullagh and Nelder[11], Lin[9]).

On the other hand, it can be easily verified that for 1-dimensional parameter θ,

$$I_\theta(g) = \Big\{E_\theta\Big(\frac{\partial g(\theta, y)}{\partial \theta}\Big)\Big\}^2 / E_\theta(g^2(\theta, y))$$

provides a compromise between maximizing concentration of the distribution of $g(\theta, y)$ and its sensitivity to θ (Birnbaum[1]). Then we call $I_\theta(g)$ the pseudo Fisher information (PFI). When θ is p-dimensional vector, in this paper we define PFI of $g(\theta, y) = (g_1(\theta, y), \ldots, g_p(\theta, y))'$ as

$$I_\theta(g) = D'(\theta)U^{-1}D(\theta), \tag{1.4}$$

where D is a $p \times p$ matrix with components $E_\theta\Big\{\frac{\partial g_i(\theta,y)}{\partial \theta_j}\Big\}$, $U = E_\theta\{g(\theta, y)g'(\theta, y)\}$.

Now suppose that the class of underlying distributions is $\mathcal{F} = \{F_\theta\}$. An estimating function g^* is said to be optimal in $\mathcal{G}$ if $g^* \in \mathcal{G}$ and if, for all $g \in \mathcal{G}$ and $F_\theta \in \mathcal{F}$, $I_\theta(g^*) \geq I_\theta(g)$ or $i_\theta(H^*) \geq i_\theta(H)$.

Conventionally, we may write $dF_\theta = f_\theta d\nu$, where f_θ is the density with respect to the measure ν on R^n. Then the true score function is $s(\theta, y) = \frac{\partial \ln f_\theta(y)}{\partial \theta}$ and the true Fisher information matrix is $I_\theta = E_\theta(s(\theta)s'(\theta))$.

In what follows, suppose that it is permitted to interchange the differentiation with respect to θ and the integration over the sample space $\mathcal{Y}$.

In sections below we obtain the upper bounds of the two classes of information matrices. Some necessary and sufficient conditions for these information matrices achieving their upper bounds are introduced. For the problem of estimating the median, some optimum estimating functions are acquired. Under some regularity conditions, an approach to carrying out the best basis function is introduced. In nonlinear regression models, an optimum estimation function is obtained. Some examples are given to illustrate our results. Finally, the concept of optimum estimating function and the methods of constructing optimum estimating function are developed in more general statistical models.

§2. The Problem of Median

2.1. Independent Sample And One-Dimensional Parameter

In this section, we first assume that $y_1, \dots, y_n$ are independently (but not necessary identically) distributed with a common median θ. From recent literature (Jung[6], Godambe and Thompso[4], Godambe[5]), we know that in order to estimate the median θ, the common estimating functions based on the independent observations $y_1, \dots, y_n$ have the form as

$$g_\phi(\theta, y) = \sum_{i=1}^{n} a_i(\theta)\phi(y_i, \theta), \tag{2.1}$$

where known functions $\phi(y_i, \theta)$, $i = 1, \cdots, n$, satisfy $E_\theta(\phi(y_i, \theta)) = 0$, and $a_i(\theta)$, $i = 1, \cdots, n$, are some arbitrary functions of θ. We call ϕ the basis function.

Let $\mathcal{G}_1(\phi)$ be a class of estimating functions as that

$$\mathcal{G}_1(\phi) = \Big\{ g_\phi : g_\phi = \sum_{i=1}^{n} a_i(\theta)\phi(y_i, \theta) \Big\}.$$

Obviously, for any $g_\phi \in \mathcal{G}_1(\phi)$, g_ϕ can be expressed as $g_\phi = H(\theta)e(\theta)$ with $H(\theta) = (a_1(\theta), \cdots, a_n(\theta))$ and $e(\theta) = (\phi(y_1, \theta), \cdots, \phi(y_n, \theta))'$. According to (1.3) and (1.4), QFI and PFI of $g_\phi \in \mathcal{G}_2(\phi)$ have the same form:

$$i_\theta(H) = I_\theta(g_\phi) = \Big\{ \sum_{i=1}^{n} a_i(\theta) E_\theta\Big(\frac{\partial \phi(y_i, \theta)}{\partial \theta} \Big) \Big\}^2 \Big/ \Big\{ \sum_{i=1}^{n} a_i^2(\theta) E_\theta(\phi^2(y_i, \theta)) \Big\}. \tag{2.2}$$

Theorem 2.1. *If observations $y_1, \cdots, y_n$ are independently distributed with common median θ, then for an arbitrary $F_\theta \in \mathcal{F}$ and $g_\phi \in \mathcal{G}_1(\phi)$, we have*

$$i_\theta(H) = I_\theta(g_\phi) \le i_\theta(H^*) = I_\theta(g_\phi^*) \le I_\theta,$$

where

$$H^*(\theta) = \Big((E_\theta(\phi^2(y_1, \theta)))^{-1} E_\theta\Big(\frac{\partial \phi(y_1, \theta)}{\partial \theta} \Big), \cdots, (E_\theta(\phi^2(y_n, \theta)))^{-1} E_\theta\Big(\frac{\partial \phi(y_n, \theta)}{\partial \theta} \Big) \Big),$$

$$g_\phi^*(\theta, y) = \sum_{i=1}^{n} (E_\theta(\phi^2(y_i, \theta)))^{-1} E_\theta\Big(\frac{\partial \phi(y_i, \theta)}{\partial \theta} \Big) \phi(y_i, \theta), \tag{2.3}$$

and $i_\theta(H^*) = I_\theta(g_\phi^*) = I_\theta$ if and only if there is a function $K(\theta)$ satisfying

$$P_\theta\{ s(\theta) = K(\theta)e(\theta) \} = 1.$$

The proof of this theorem is presented in Appendix below.

Theorem 2.1 shows that the optimum estimating function in $\mathcal{G}_1(\phi)$ is $g_\phi^*(\theta, y)$ as defined by (2.3). The theorem also implies that two classes of information of estimating function $g(\theta, y)$ equal to their upper bound I_θ if and only if true score function $s(\theta)$ belongs to $\mathcal{G}_1(\phi)$ with probability 1. If ϕ can be expressed as $\phi(y_i, \theta) = c(\theta)T(y_i) + b(\theta)$ for some functions $c(\theta)$ and $b(\theta)$, this is equivalent to

$$f_\theta(y) = C(\theta) \exp\Big\{ \sum_{i=1}^{n} Q_i(\theta) T(y_i) \Big\} h(y) \text{ a.s.}$$

for some functions $C(\theta)$, $Q_i(\theta)$, $i = 1, \cdots, n$, and $h(y)$, i.e. f_θ does belong to the family of exponential distribution. The results above lead to the following corollary.

Corollary 2.1. *Assume that observations $y_1, \cdots, y_n$ are independently distributed with median θ.*

(1) *If the score function has the form $s(\theta) = K(\theta)e(\theta)$ a.s. for some function $K(\theta)$ and vector $e(\theta) = (\phi(y_1, \theta), \cdots, \phi(y_n, \theta))'$, then ϕ is the best basis function, i.e. $I_\theta(g^*_\phi) = I_\theta$, if and only if ϕ is unbiased.*

(2) *If the density function has the form $f_\theta(y) = C(\theta) \exp\Big\{ \sum\limits_{i=1}^{n} Q_i(\theta)T(y_i)\Big\}h(y)$ a.s. for some functions $C(\theta)$, $Q_i(\theta)$, $i = 1, \cdots, n$, and $h(y)$, and ϕ can be expressed as $\phi(y_i, \theta) = c(\theta)T(y_i) + b(\theta)$ for some functions $c(\theta)$ and $b(\theta)$, then ϕ is the best basis function, i.e. $I_\theta(g^*_\phi) = I_\theta$, if and only if ϕ is unbiased.*

From this corollary, we are able to construct the best basis function if the score function $s(\theta)$ or density function $f_\theta(y)$ and basis function ϕ have above special forms. In usual situation, however, the score function or density function is unknown or is not framed as in Corollary 1. So, it is difficult to determine which basis function to use. In Section 4, using the corollary, we give some examples to compare some basis functions.

From the proof of Theorem 2.1, on the other hand, we can see that to obtain some results in Theorem 1, the condition such as $g_\phi \in \mathcal{G}_1(\phi)$ is not necessary, i.e. we have the following corollary.

Corollary 2.2. *If observations $y_1, \cdots, y_n$ are independently distributed with median θ, then for an arbitrary $F_\theta \in \mathcal{F}$ and $g \in \mathcal{G}$, we have*

$$I_\theta(g) \le I_\theta,$$

where $\mathcal{G} = \{g(\theta, y) : E_\theta(g) = 0\}$, and $I_\theta(g) = I_\theta$ if and only if there is a function $K(\theta)$ satisfying $P_\theta\{s(\theta) = K(\theta)g(\theta, y)\} = 1$.

In the examples below, the distribution function F_θ is assumed to be continuous. To obtain the optimum estimating function, a suitable basis function ϕ is important. We initially consider the following basis function

$$\phi(y_i, \theta) = \text{sign}(y_i - \theta). \tag{2.4}$$

According to common definition of derivative, however, $\frac{\partial \phi}{\partial \theta}$ can not be defined on some points. In this case, following Godambe and Thompson[4], we define

$$E_\theta\Big(\frac{\partial \phi(y_i, \theta)}{\partial \theta}\Big) = \lim_{\varepsilon \to 0} E_\theta\left((\phi(y_i, \theta + \varepsilon) - \phi(y_i, \theta))/\varepsilon\right).$$

It can be verified that for basis function (2.4),

$$-1/2E_\theta\Big(\frac{\partial \phi(y_i, \theta)}{\partial \theta}\Big) = \lim_{\varepsilon \to 0} P_\theta(\theta < y_i < \theta + \varepsilon)/\varepsilon \triangleq f_i(\theta),$$

where $f_i(\theta) > 0$ is the density of y_i at its median. Thus, in the case, the optimum estimating function $g^*_\phi = \sum\limits_{i=1}^{n} f_i(\theta)\phi(y_i, \theta)$ ignoring a constant -2. If observations $y_1, \cdots, y_n$, in addition to being independent, are identically distributed, then $f_1(\theta) = \cdots = f_n(\theta)$. As a result the optimum estimating function g^*_ϕ is equivalent to $g^*_\phi = \sum\limits_{i=1}^{n} \phi(y_i, \theta)$, and the solution $\hat{\theta}$ of the

equation $g_\phi^* = 0$ can be easily seen to be the ordinary median of the observations $y_1, \cdots, y_n$.

According to empirical, there are some other methods to choose basis function. For example (Jung[6]), a common choice is

$$\phi(y_i, \theta) = I(y_i \geq \theta) - 1/2, \tag{2.5}$$

where $I(\cdot)$ is an indicator function. Similar to the result discussed above, in this situation, the optimum estimating function $g_\phi^* = \sum_{i=1}^n f_i(\theta)\phi(y_i, \theta)$, where $f_i(\theta) > 0$ is also the density of y_i at its median. It can be verified that, whether the basis function ϕ is defined by (2.4) or by (2.5) or more generally by

$$\phi(y_i, \theta) = \begin{cases} -c, & \text{if } y_i < \theta, \\ +c, & \text{if } y_i \geq \theta, c > 0, \end{cases}$$

the values of QFI and PFI g_ϕ^* are equal to the same value: $4\sum\limits_{i=1}^n f_i^2(\theta)$. In Section 4, we will give examples to introduce the condition under which these basis functions are the best basis functions.

From the examples, we also see that in the usual situation, the values of f_i at their common median θ must be known as explicit function of θ, up to a constant multiplier.

2.2. Dependent Sample and Multidimensional Parameter

In this section we first suppose observations $y_1, \cdots, y_n$ are distributed independently with respective median $\theta_1 + \theta_2 x_1, \cdots, \theta_1 + \theta_2 x_n$, where θ_1 and θ_2 are unknown parameters and $x_1, \cdots, x_n$ are fixed design covariates. If the basis function is defined as

$$\phi(y_i, \theta_1, \theta_2) = \text{sign}(y_i - (\theta_1 + \theta_2 x_i)), \tag{2.6}$$

then, similar to the results above, the jointly optimum estimating function $g_\phi^* = (g_{1\phi}^*, g_{2\phi}^*)'$ for estimating parameters θ_1 and θ_2 is given by

$$g_{1\phi}^* = \sum_{i=1}^n \phi(y_i, \theta_1, \theta_2) f_i(\theta_1 + \theta_2 x_i) \quad \text{and} \quad g_{2\phi}^* = \sum_{i=1}^n \phi(y_i, \theta_1, \theta_2) f_i(\theta_1 + \theta_2 x_i) x_i,$$

where $f_i(\theta_1 + \theta_2 x_i) > 0$ is the density of y_i at the point $\theta_1 + \theta_2 x_i$.

More generally, consider the following multi-parameter median regression model[6]:

$$\begin{cases} \text{med}(y) = \theta(\beta) = (\theta_1(\beta), \cdots, \theta_n(\beta))' \\ \text{pdf of } y_i \text{ at } \theta_i = \frac{1}{\tau}\gamma(\theta_i), \end{cases} \tag{2.7}$$

where $y = (y_1, \cdots, y_n)', y_1, \cdots, y_n$ may be independent or not, $\tau > 0$ is a scale parameter, γ is a known function relating the median to the degree of dispersion, $\beta = (\beta_1, \cdots, \beta_p)'$ is the vector of regression parameters.

Let $\phi_i(y_i, \beta)$ be basis function satisfying

$$E_\beta(\phi_i(y_i, \beta)) = 0, \quad e(\beta) = (\phi_1(y_1, \beta), \cdots, \phi_n(y_n, \beta))',$$

$H(\beta)$ be a $p \times n$ matrix and $\mathcal{G}_1(\phi)$ be a class of estimating functions as that

$$\mathcal{G}_1(\phi) = \{g_\phi : \, g_\phi = H(\beta)e(\beta)\}.$$

According to (1.3), in this situation, QFI of an arbitrary unbiased estimating function $g_\phi = H(\beta)e(\beta)$ is defined as

$$i_\beta(H) = \Delta'(\beta)H'(\beta)(H(\beta)\Lambda(\beta)H'(\beta))^{-1}H(\beta)\Delta(\beta), \tag{2.8}$$

where Δ is an $n \times p$ matrix with components $E_\beta\{\partial\phi_i(y_i,\beta)/\partial\beta_j\}$ and Λ is an $n \times n$ matrix with components $E_\beta\{\phi_i(y_i,\beta)\phi_j(y_j,\beta)\}$. According to (1.4), in this situation, we define PFI of $g_\phi = H(\beta)e(\beta)$ as

$$I_\beta(g_\phi) = D'(\beta)U^{-1}D(\beta) = i_\beta(H). \tag{2.9}$$

Theorem 2.2. *For an arbitrary $F_\theta \in \mathcal{F}$ and $g_\phi \in \mathcal{G}_1(\phi)$, we have*

$$i_\beta(H) = I_\beta(g_\phi) \le i_\beta(\Delta'\Lambda^{-1}) = I_\beta(g_\phi^*) \le I_\beta,$$

where

$$g_\phi^* = \Delta'(\beta)\Lambda^{-1}(\beta)e(\beta), \tag{2.10}$$

and $i_\beta(\Delta'\Lambda^{-1}) = I_\beta(g_\phi^*) = I_\beta$ if and only if there is a matrix $K(\beta)$ satisfying

$$P_\beta\{s(\beta) = K(\beta)e(\beta))\} = 1.$$

The proof of this theorem is similar to that of Theorem 3.1 presented below.

Theorem 2.2 shows that the optimum estimating function in $\mathcal{G}_1(\phi)$ is g_ϕ^* as defined by (2.10). If the basis function $\phi_i(y_i,\beta)$ is chosen as $I(y_i \ge \theta_i) - \frac{1}{2}$, then, the optimum estimating function g_ϕ^* has the same form as defined as Jung[6]. Under moderate assumptions, the estimate of β obtained by equation $g_\phi^* = 0$ is consistent and has asymptotically normal distribution[6].

The theorem also implies that QFI and PFI of estimating function g_ϕ equal to their upper bound I_β if and only if true score function $s(\beta)$ belongs to $\mathcal{G}_1(\phi)$, with probability 1. Then we get the following corollary.

Corollary 2.3. (1) *In model* (2.7), *if the score function has the form $s(\theta) = K(\theta)e(\theta)$ a.s. for some matrix $K(\theta)$ and vector $e(\theta) = (\phi_1(y_1,\theta), \cdots, \phi_n(y_n,\theta))'$, then ϕ_i are the best basis functions, i.e. $I_\theta(g_\phi^*) = I_\theta$, if and only if ϕ_i are unbiased.*

(2) *In model* (2.7), *if the density of $(y_1, \cdots, y_n)$ has the following form*

$$f_\theta(y) = C(\beta)\exp\Big\{\sum_{i=1}^n Q_i(\beta)T_i(y_i)\Big\}h(y) \ \text{a.s.}$$

for some functions $C(\beta)$, $Q_i(\beta)$, and $h(y)$, and ϕ_i can be expressed as $\phi_i(y_i,\beta) = c_i(\beta)T_i(y_i) + b_i(\beta)$ for some functions $c_i(\beta)$, $T_i(y_i)$ and $b_i(\beta)$, then ϕ_i are the best basis functions, i.e. $I_\theta(g_\phi^*) = I_\theta$, if and only if ϕ_i are all unbiased.

From this corollary, we are able to construct the set of the best basis functions if the density function $f_\theta(y)$ and the basis functions have above special forms. In usual situation, however, the density function is unknown or is not framed as in Corollary 2.3. So, it is difficult to determine which set of basis functions to use.

§3. Nonlinear Regression Model

In this section, we assume that $n \times 1$ observation y has mean $\mu(\theta)$ and covariance matrix $\sigma^2 V(\theta)$, both being known functions of the p-dimensional parameter θ and $V(\theta)$ being a

positive definite matrix. In other words, we have the following nonlinear regression model

$$\begin{cases} y = \mu(\theta) + \varepsilon, \\ E_\theta(\varepsilon) = 0, \qquad Var_\theta(\varepsilon) = \sigma^2 V(\theta). \end{cases} \tag{3.1}$$

Let

$$\mathcal{G}_2 = \{g : g = \sigma^{-2} H(\theta) e(\theta), e(\theta) = y - \mu(\theta)\}.$$

In this case, for an arbitrary estimating function $g \in \mathcal{G}_2$, QFI defined in (1.3) and PFI defined in (1.4) are equal to each other, i.e.

$$i_\theta(H) = I_\theta(g) = \sigma^{-2} \dot{\mu}'(\theta) H'(\theta) (H(\theta) V(\theta) H'(\theta))^{-1} H(\theta) \dot{\mu}(\theta). \tag{3.2}$$

Theorem 3.1. *For an arbitrary $F_\theta \in \mathcal{F}$ and $g \in \mathcal{G}_2$, we have*

$$i_\theta(H) = I_\theta(g) \le i_\theta(\dot{\mu}' V^{-1}) = I_\theta(q) \le I_\theta,$$

where q is defined by (1.1)*, and $i_\theta(\dot{\mu}' V^{-1}) = I_\theta(q) = I_\theta$ if and only if there is a matrix $K(\theta)$ satisfying*

$$P_\theta\{s(\theta) = \sigma^{-2} K(\theta)(y - \mu(\theta))\} = 1.$$

The proof of this theorem is given in Appendix below.

Theorem 3.1 shows that under the criterions of QFI and PFI, the quasi score function $q(\theta, y)$ defined by (1.1) is an optimum estimating function in $\mathcal{G}_2$. The theorem also implies that QFI and PFI of quasi score function $q(\theta, y)$ equal to their upper bound I_θ if and only if true score function $s(\theta)$ belongs to $\mathcal{G}_2$ with probability 1, this is equivalent to that the observation y, with probability 1, comes from a family of exponential distribution, i.e.

$$f_\theta(y) = C(\theta, \sigma^2) \exp\{\sigma^{-2} Q(\theta) y\} h(y) \text{ a.s.}$$

for some functions $C(\theta, \sigma^2)$, $Q(\theta)$ and $h(y)$.

The quasi score method has been investigated in some other respects by many statisticians such as Wedderburn[12], Godambe and Heyde[2], Li and McCullagh[7] and so on. According to the theorem above, we obtain again its optimality based on the two classes of information.

§4. Extensions

The concepts and the methods as discussed above can be developed in more general statistical models. Assume that we are able to determine the set of basis functions $\phi_1(y_1, \theta), \cdots, \phi_n(y_n, \theta)$. All functions $\phi_i(\cdot, \cdot)$ may be the same function as presented in Section 2 or not, may be the residual $e_i(\theta) = y_i - \mu_i(\theta)$ as expressed in Section 3 or not. The essential characteristic of the basis functions is that the basis functions are unbiased, i.e. $E_\theta(\phi_i(y_i, \theta)) = 0$ for all θ. The parameter θ may be the median as discussed in Section 2 or not, may be the regression parameter as in Section 2 or not. In fact, θ is an arbitrary parameter in a statistical model. The observations, $y_1, \cdots, y_n$, may be independently distributed or not, may be identically distributed or not. Under these usual assumptions, to estimate θ, a p-dimensional vector of parameters, the class of estimating functions is defined as

$$\mathcal{G}(\phi) = \{g_\phi : g_\phi = H(\theta) e(\theta), H(\theta) \text{ is a } p \times n \text{ matrix}, \ e(\theta) = (\phi_1(y_1, \theta), \cdots, \phi_n(y_n, \theta))'\}.$$

According to Theorem 2.2, in this situation, we have the following results:

For an arbitrary $F_\theta \in \mathcal{F}$ and $g_\phi \in \mathcal{G}(\phi)$, the QFI and PFI satisfy

$$i_\theta(H) = I_\theta(g_\phi) \le i_\theta(\Delta'\Lambda^{-1}) = I_\theta(g_\phi^*) \le I_\theta,$$

where

$$g_\phi^* = \Delta'(\theta)\Lambda^{-1}(\theta)e(\theta),$$

Δ and Λ are defined in Section 2, and

$$i_\theta(\Delta'\Lambda^{-1}) = I_\theta(g_\phi^*) = I_\theta$$

if and only if there is a matrix $K(\theta)$ satisfying

$$P_\theta\{s(\theta) = K(\theta)e(\theta)\} = 1.$$

Similar to the results of Jung[6], under some regularity conditions, the estimator $\hat{\beta}$, being the root of the equation $g_\phi^* = 0$, has the following properties:

$$\hat{\beta} \to \beta_0 \text{ a.s. and } \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma),$$

where

$$\Sigma = \lim_{n\to\infty}(n)^{-1} i_\theta(\Delta'\Lambda^{-1}).$$

On the other hand, choosing some suitable basis functions is also important because different basis functions perhaps have different values of QFI and PFI. By the result above or Corollary 2.1 or Corollary 2.3, we sometimes can get the best basis functions. See the following examples.

In two-parameter double exponential family of distribution, the density function

$$f_\theta(y_1, \cdots, y_n) = \prod_{i=1}^{n}(2\lambda_i)^{-1} \exp\{-|y_i - \theta|/\lambda_i\} \quad \text{a.s.},$$

where θ is the median and λ_i is a scale parameter. If the basis function is chosen as $\tilde{\phi}(y_i, \theta) = y_i - \theta$, then the value of QFI and PFI is $\sum\limits_{i=1}^{n}(4\lambda_i^2)^{-1}$. It can be verified that the score function

$$s(\theta) = -\sum_{i=1}^{n}(\lambda_i)^{-1}\phi(y_i, \theta) \text{ a.s.},$$

where ϕ is presented as (2.4). If the basis function is chosen as (2.4), then, the value of QFI and PFI is $\sum\limits_{i=1}^{n}\lambda_i^{-2}$, which is just the Fisher information. Thus basis function (2.4) is the best basis function and then is better than $\tilde{\phi}(y_i, \theta) = y_i - \theta$.

Assume $y_1, \cdots, y_n$ are $i.i.d$ and come from $N(\mu, \sigma^2)$. The score function for μ is

$$s(\mu) = \sigma^{-2}\sum_{i=1}^{n}(y_i - \mu).$$

If the basis function is chosen as $\tilde{\phi}(y_i, \mu) = y_i - \mu$, then the value of QFI and PFI is $n\sigma^{-2}$, which is equal to Fisher information. If the basis function is chosen as (2.4), then value of QFI and PFI is $2n(\pi\sigma)^{-2}$, which is smaller than $n\sigma^{-2}$. In this case, $\tilde{\phi}(y_i, \mu) = y_i - \mu$ is the best basis function in the class of unbiased basis functions.

Generally, assume that y has a continuous density function $f(|y-\theta|/\lambda)$ and $x=0$ is the symmetric center of function $f(x)$. Let $y_1,\cdots,y_n$ be independently and identically distributed observations of y. If the basis function is chosen as $\tilde{\phi}(y_i,\theta)=y_i-\theta$, then the value of QFI and PFI is n/λ^2, where λ^2 is just the variance of y. If the basis function is chosen as (2.4), then the value of QFI and PFI is $1/(4f^2(0))$. When we know the values of λ^2 and $f(0)$ or $\lambda^2/f^2(0)$ (not necessary to know the distribution of y), we can conclude which is better.

§5. Appendix

Proof of Theorem 2.1. Obviously,

$$\begin{aligned} i_\theta(H)=I_\theta(g_\phi)&=\frac{\Big\{\sum\limits_{i=1}^n a_i(\theta)E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)\Big\}^2}{\sum\limits_{i=1}^n a_i^2(\theta)E_\theta(\phi^2(y_i,\theta))}\\ &=\frac{\Big\{\sum\limits_{i=1}^n[a_i(\theta)(E_\theta(\phi^2(y_i,\theta)))^{1/2}]\Big[(E_\theta(\phi^2(y_i,\theta)))^{-1/2}E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)\Big]\Big\}^2}{\sum\limits_{i=1}^n a_i^2(\theta)E_\theta(\phi^2(y_i,\theta))}\\ &\le\sum_{i=1}^n(E_\theta(\phi^2(y_i,\theta)))^{-1}\Big[E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)\Big]^2, \end{aligned}$$

$$\begin{aligned} I_\theta(g_\phi^*)&=\frac{\Big\{\sum\limits_{i=1}^n(E_\theta(\phi^2(y_i,\theta)))^{-1}\Big(E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)\Big)^2\Big\}^2}{\sum\limits_{i=1}^n(E_\theta(\phi^2(y_i,\theta)))^{-1}\Big(E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)^2}\\ &=\sum_{i=1}^n(E_\theta(\phi^2(y_i,\theta)))^{-1}\Big[E_\theta\Big(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big)\Big]^2. \end{aligned}$$

Then $I_\theta(g_\phi)\le I_\theta(g_\phi^*)$.

On the other hand, for any unbiased estimating function g, from $E_\theta(g(\theta,y))=0$, we get $\frac{\partial E_\theta(g(\theta,y))}{\partial\theta}=0$. If the differential and the integration are interchanged, then

$$E_\theta\{g(\theta,y)s'(\theta)\}=-E_\theta\Big\{\frac{\partial g(\theta,y)}{\partial\theta}\Big\}\triangleq -D. \tag{5.1}$$

Let $W(\theta)=s(\theta)+D'U^{-1}(\theta)g(\theta,y)$ and $U(\theta)=E_\theta(g^2(\theta,y))$. Obviously, $E_\theta(W(\theta))=0$. From (5.1), we obtain

$$\begin{aligned} 0\le\mathrm{Cov}_\theta(W)&=E_\theta\{(s+D'U^{-1}g)(s+D'U^{-1}g)'\}\\ &=E_\theta(ss')+D'U^{-1}E_\theta(gs')+\{E_\theta(sg')\}'U^{-1}D+D'U^{-1}D\\ &=F_\theta-D'U^{-1}D. \end{aligned}$$

From the result above, we can see that $I_\theta\ge D'U^{-1}D=I_\theta(g)$ and $\mathrm{Cov}_\theta(W)=0$, i.e. $I_\theta=I_\theta(g)$, if and only if there is $K(\theta)$ satisfying $s(\theta)=K(\theta)g(\theta,y)$ with probability 1. Thus the theorem is completely proved.

Proof of Theorem 3.1. It is clear from (1.2) and (3.2) that for $g \in \mathcal{G}_1$,

$$\begin{aligned} i_\theta - i_\theta(H) &= i_\theta - I_\theta(g) = \sigma^{-2}\dot{\mu}'V^{-1}\dot{\mu} - \sigma^{-2}\dot{\mu}'H'(HVH')^{-1}H\dot{\mu} \\ &= \sigma^{-2}\dot{\mu}'V^{-1/2}\left(I - V^{1/2}H'(HVH')^{-1}HV^{1/2}\right)V^{-1/2}\dot{\mu}. \end{aligned}$$

Since $M \triangleq V^{1/2}H'(HVH')^{-1}HV^{1/2}$ satisfies $M' = M$ and $M^2 = M$, the difference above is non-negative. Then we get

$$i_\theta(H) = I_\theta(g) \le i_\theta.$$

On the other hand, since $E_\theta(y - \mu(\theta)) = 0$, $\frac{\partial E_\theta(y-\mu(\theta))}{\partial\theta} = 0$. If the differential and the integration are interchanged, then

$$E_\theta\{(y - \mu(\theta))s'(\theta)\} = -E_\theta\Big\{\frac{\partial(y - \mu(\theta))}{\partial\theta}\Big\} = \dot{\mu}(\theta). \tag{5.2}$$

Let $W(\theta) = s(\theta) - \sigma^{-2}\dot{\mu}'(\theta)V^{-1}(\theta)(y - \mu(\theta))$. Obviously, $E_\theta(W(\theta)) = 0$. From (5.2), we obtain

$$\begin{aligned} 0 \le \mathrm{Cov}_\theta(W) &= E_\theta\{(s - \sigma^{-2}\dot{\mu}'V^{-1}(y - \mu))(s - \sigma^{-2}\dot{\mu}'V^{-1}(y - \mu))'\} \\ &= E_\theta(ss') - \sigma^{-2}\dot{\mu}'V^{-1}E_\theta\{(y - \mu)s'\} - \sigma^{-2}E_\theta\{s(y - \mu)'\}V^{-1}\dot{\mu} + \sigma^{-2}\dot{\mu}'V^{-1}\dot{\mu} \\ &= F_\theta - \sigma^{-2}\dot{\mu}'V^{-1}\dot{\mu}. \end{aligned}$$

From the result above, we can see that $F_\theta \ge \sigma^{-2}\dot{\mu}'V^{-1}\dot{\mu}$ and $\mathrm{Cov}_\theta(W) = 0$, i.e. $F_\theta = \sigma^{-2}\dot{\mu}'V^{-1}\dot{\mu}$, if and only if there is a matrix $K(\theta)$ satisfying $s(\theta) = \sigma^{-2}K(\theta)(y - \mu(\theta))$ with probability 1. Thus the theorem is completely proved.

Acknowledgement. The author would like to thank Prof. R. C. Zhang for his help.

References

[1] Birnbaum, A., Statistical theory for logistic metal test models with a prior distribution of ability, *Journal of Mathematical Psychology*, **37**(1972), 29–51.
[2] Godambe, V. P. & Heyde, C. D., Quasi-likelihood and optimal estimation, *Interal Statist Rev.*, **55**(1987), 231–244.
[3] Godambe, V. P. & Thompson, M. E., An extension of quasi-likelihood estimation, *J. Statist. Plan. Inference*, **22**(1989), 137–152.
[4] Godambe, V. P. & Thompson, M. E., Robust estimation through equations, *Biometrika*. **71**(1984), 115–125.
[5] Godambe, V. P., Estimation of method: quasi-likelihood and optimum estimating function, Working Paper, 2001-04, Department of Statistics and Actuarial Science University of Waterloo 2001.
[6] Jung, S., Quasi-Likelihood for Median Regression Models, *JASA*, **91**(1996), 251–257.
[7] Li, B.& McCullagh, P., Potential functions and conservative estimating functions, *Annals of Statistics*. **22**(1994), 341–356.
[8] Lin, L., Conservative Estimating Function in Nonlinear Model with Aggregated Data. *Mathematica Acta Sinientia,* **20:**3(2000), 335–340.
[9] Lin, L., Some Properties for Quasi likelihood Estimate in Nonlinear Regression Models. *Acta Mathematicae Applicatae Sinica.* **22:**2(199), 307–310.
[10] McCullagh, P., Quasi-Likelihood Functions. *Annals of Statistics*. **11**(1983), 59–67.
[11] McCullagh, P. & Nelder, J. A., Generalized linear models, Second Edition. London, *Chmpman and Hall*:, 1989, 323–339.
[12] Wedderburn, R. W. M., Quasi-likelihood functions, generalized linear models, and the Guess-Newton method, *Biometrika,* **61**(1974), 439–447.